

UNCLASSIFIED

AD NUMBER

ADB300928

LIMITATION CHANGES

TO:

Approved for public release; distribution is unlimited.

FROM:

Distribution authorized to U.S. Gov't. agencies and their contractors;  
Administrative/Operational Use; 2004. Other requests shall be referred to Defense Technical Information Center, 8725 John J. Kingman Rd., Ft. Belvoir, VA 22060-6218.

AUTHORITY

DTIC ltr, 15 Mar 2010

THIS PAGE IS UNCLASSIFIED

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 1995	3. REPORT TYPE AND DATES COVERED Final 1995		
4. TITLE AND SUBTITLE  Gray Literature		5. FUNDING NUMBERS		
6. AUTHOR(S) R. Paul Ryan and Mason H. Soule				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Battelle Memorial Institute 505 King Avenue Columbus, Ohio 43201		8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Defense Technical Information Center 8725 John J. Kingman Road Ft. Belvoir, VA 22060-6218		10. SPONSORING/MONITORING AGENCY REPORT NUMBER		
11. SUPPLEMENTARY NOTES Presentation to the Battelle/DTIC "summit" held August 10, 1995. Copyright Battelle 1995. U.S. Government Joint Work (unlimited government copyrights).				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Distribution limited to US Government and contractors; Administrative/Operational Use, 2004. Other requests for this document shall be referred to the Defense Technical Information Center, 8725 John J. Kingman Rd, Ft. Belvoir, VA 22060-6218		12b. DISTRIBUTION CODE  C		
12. ABSTRACT (Maximum 200 words)  This paper provides a brief introduction to the concept of gray literature, successively discussing issues of scope, problems and benefits, as well as Federal Government activities to address and hopefully solve some of the problems raised. The paper concludes by commenting on the roles that information technologies can play in improving the capability, particularly in the Intelligence Community, to deal with tricky type of information.				
SUBJECT TERMS Information management, Gray literature			15. NUMBER OF PAGES 6	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT  UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE  UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT  UNCLASSIFIED	20. LIMITATION OF ABSTRACT  UNCLASSIFIED	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)  
Prescribed by ANSI Std. Z39-18  
298-102

## **GRAY LITERATURE**

Presentation to  
The Battelle/DTIC "summit"  
August 10, 1995

Mason H. Soule  
BATTELLE  
soule@battelle.org  
614/424-6191

R. Paul Ryan  
DEFENSE TECHNICAL INFORMATION CENTER  
ryan@dtic.mil  
703/767-9200

## **OVERVIEW**

To respond to changes in international events, the Intelligence Community (IC) must formulate new approaches to provide timely, dependable, and actionable intelligence information to U.S. policy makers and warfighters. In the open source domain, the IC has long taken advantage of the large, worldwide publishing system -- that is, "normal" bookselling channels -- to identify and acquire journals, serials, newspapers, books, databases, and other open source materials of intelligence value. However, one of the IC's major open source challenges is to acquire and utilize the growing body of gray information of intelligence value. "Gray" information or gray literature<sup>1</sup> is principally distinguished as being outside of these normal bookselling channels, which makes it relatively more difficult to identify and acquire than other open source literature. In order to improve its coverage of this type of information, the IC must devise ways to simulate the awareness and document supply functions that the broader bookselling system provides, while recognizing that the resources do not exist to create a duplicate system for gray information.

This paper provides a brief introduction to the concept of gray literature, successively discussing issues of scope, problems, and benefits, as well as recent Federal Government activities to address and hopefully solve some of the problems raised. Since this summit focuses on information technologies, we conclude by commenting on the role that these technologies can play in improving our capability to deal with this tricky type of information.

### **(1) SCOPE**

The library science community broadly distinguishes among "white" or published literature (books and journals, mainly having ISBN or ISSN numbers), "ephemeral" literature (items of very short-lived interest such as printed airline schedules), and "gray" literature, that literature which falls between these extremes. The IC, as part of its program to improve its exploitation of open source material, developed a definition of gray literature to suit its particular needs. It was generated and adopted by the U.S. Government's Interagency Gray Literature Working Group (IGLWG) to support the group's charter of preparing a Gray Information Functional Plan for member agencies both inside and outside the IC. Building on a Library of Congress definition,

---

<sup>1</sup> We use the phrases gray literature and gray information interchangeably in this paper. The former traditionally denotes hardcopy books and journals; the latter implicitly includes other types of media.

the IGLWG distinguished gray literature from other open sources:

"Gray literature is foreign or domestic open source material that usually is available through specialized channels and may not enter normal channels or systems of publication, distribution, bibliographic control, or acquisition by booksellers or subscription agents."<sup>1</sup>

There are many other interpretations of what constitutes gray literature. It remains difficult to define because the boundaries between it and other open source information types vary by user group, and are fuzzy and variable in time and space. But, it is important to those tasked with acquiring this information to have as little ambiguity as possible. Accordingly, the IGLWG included a discussion of the types of information conventionally considered gray. These include, but are not limited to:

- |                      |                    |
|----------------------|--------------------|
| -academic papers     | -preprints         |
| -committee reports   | -proceedings       |
| -conference papers   | -research reports  |
| -corporate documents | -standards         |
| -discussion papers   | -technical reports |
| -dissertations       | -theses            |
| -government reports  | -trade literature  |
| -house journals      | -translations      |
| -market surveys      | -trip reports      |
| -newsletters         | -working papers.   |

Also as an aid to collectors the IGLWG noted the major kinds of organizations that produce gray literature:

"Organizations that typically generate the largest quantities of gray literature include: research establishments (laboratories and institutes); national governments; private publishers (pressure groups/political parties); corporations; trade associations/unions; think tanks; and academia."

These clarifications are important because they provide guidance to collectors who need to know what kind of information their customers want to obtain. But in summary, the key point to make regarding scope is the following: Since gray literature is not well-covered by conventional book trade channels, it is relatively more difficult to identify, acquire, process, and access than conventional open source literature. Hence, if we desire to use gray literature as a source of information, we must be prepared to accept a greater expenditure of resources to collect and process this information in comparison to other open source material.

---

<sup>1</sup> Interagency Gray Literature Working Group (IGLWG), "Gray Information Functional Plan (GIFP)," 18 January, 1995.

## **(2) PROBLEMS WITH GRAY LITERATURE**

There are numerous difficulties in using gray information and here we mention only a few major ones. First and foremost, and as the previous section concludes, gray literature is difficult to search for, identify, and acquire. This puts significantly more burden on the traditional "collection" stage of the intelligence cycle. For example, the only way to learn about or to acquire some trade literature and unpublished conference papers is to attend the functions at which they are made available. However, our networks for identifying events before the fact are poorly developed since again they are not often covered by traditional channels.

Second, open source already suffers from the problem of a low signal-to-noise ratio, i.e., there often are very few nuggets to be sifted from this large body of information. The problem is exacerbated in the gray literature domain because thousands of organizations generate literature while only a fraction of these producers and their products are of interest to the IC. This situation is worsening daily as the availability of information from myriad Internet sites increases.

Third, gray literature is more difficult to process than other open source types because of its predominantly nonstandard formats. Product brochures, for example, rarely provide adequate information to allow them to be catalogued or retrieved easily. Author, title, place and date of publication, and publisher often are lacking from other gray literature types, as well. In addition, much gray information remains available only in hard-copy. Although this is changing as Internet distribution expands, the absence of standards will make it difficult to find information on this electronic forum with other than direct character matches.

Fourth, gray literature varies radically in quality since it often is unrefereed. Integrity is an issue with Internet data, as well, since electronic data are easy to alter.

Finally, foreign gray materials, which are the main interest of the IC, are much less likely to be in English than their white literature counterparts. This places additional burden on the processing system which needs human or machine translators to translate the material for the user.

## **(3) IMPORTANCE OF GRAY LITERATURE**

Key reasons for recognizing gray literature as a subset of open source lie not in the problems it generates but in the value it provides. The Intelligence Community's interest in open source stems from its potential to contain information of intelligence value and which may be obtained cheaper, faster, and at a lower level of classification than information acquired through non-open source intelligence collection channels. As a subset of open source, gray information has other attributes, as well. First, it can provide information that often is unavailable in published open sources. Many brochures and the information they contain never will appear in a published version.

Second, gray literature often is available on a more timely basis than conventional literature. Conference papers, as an example, are available long before any follow-on, published article will appear, yet the information content of the two versions may not be

significantly different.

Third, it can corroborate important assertions found in other sources, which is always paramount in intelligence analysis.

Fourth, it may have a concise, focused, and detailed content. This is particularly true of technical reports or unofficial government documents, whose information content will be greatly reduced in the published form.

And fifth, it is becoming a common means of information exchange, particularly as personal publishing software improves and Internet expands.

For all these reasons, gray literature must be part of an overall awareness strategy that requires a thorough search be made of all open sources to answer intelligence questions.

#### **(4) FEDERAL GOVERNMENT INITIATIVES TO ADDRESS GRAY LITERATURE**

The successful exploitation of gray literature requires coordination and sharing of knowledge to reduce average unit costs of gray literature to all involved parties. Recognizing this, in February 1993, the Director of the Foreign Broadcast Information Service (D/FBIS) asked for IC and non-IC participation to develop a functional plan for acquiring, processing, and disseminating gray information throughout the IC. In response to that call, representatives from the Armed Forces Medical Intelligence Center (AFMIC), Army Materiel Command (AMC), Community Open Source Program Office (COSPO), Defense Intelligence Agency (DIA), Defense Technical Information Center (DTIC), Department of the Army (DA), Department of Energy (DOE), Library of Congress, National Air Intelligence Center (NAIC), National Ground Intelligence Center (NGIC), National Security Agency (NSA), National Aeronautics and Space Administration (NASA), and National Technical Information Center (NTIS), met with FBIS and other CIA Headquarters components to form the Interagency Gray Literature Working Group (IGLWG).

The Gray Information Functional Plan created by this group was approved by D/FBIS and D/COSPO and released in January, 1995. It addresses issues of acquisition, processing, and dissemination. The plan lists ten key findings that represent the basic precepts for action to make gray information available to help fulfill intelligence needs in the 1990s:

- \* develop an AWARENESS of the availability of gray information within or to organizations as a means to minimize acquisition and processing costs to other agencies
- \* improve IC COOPERATION AND COORDINATION with key non-IC members that currently collect the majority of gray information materials to maximize its availability and to reduce duplicate acquisition
- \* implement NEEDS DRIVEN ACQUISITION of gray information to better manage acquisition costs

- \* identify and assign AREAS OF RESPONSIBILITY among Government Agencies to cover myriad subject areas on a worldwide basis with a dwindling resource base
- \* provide TIMELY NOTIFICATION of the availability of relevant gray information
- \* utilize a TIERED PROCESSING APPROACH--driven by user requests--to minimize total processing costs
- \* institute LONG-TERM STORAGE AND RETRIEVAL systems
- \* leverage EXISTING US GOVERNMENT CAPABILITIES FOR DEPOSIT AND DISTRIBUTION to reduce infrastructure costs
- \* incorporate advancements in ENABLING EXPLOITATION TECHNOLOGIES (principally optical character recognition, machine translation, and machine-aided indexing) to improve processing
- \* employ credible MANAGEMENT METRICS to ensure the continuing effectiveness and efficiency of systems and processes.

The IGLWG believes that gray information can be better harnessed to satisfy intelligence needs if we can carefully implement these basic ideas among acquisition agencies and wisely use existing Government processing and distribution functions.

Members of the IGLWG have been involved in other ways to improve access to gray literature, both as individual organizations and as part of multiagency ventures. Over the past five years, first NASA and later FBIS, NAIC, DTIC, and COSPO have co-sponsored the annual International Acquisitions Workshop to share knowledge of gray and other open source collections and acquisition methodologies among Government information specialists. Under a COSPO-funded project, NAIC developed a Gray Literature On-Line Catalog (GLOC), a database residing on the CIRC system which describes the gray literature holdings of many Government and non-Government information centers. FBIS has developed a Gray Literature Tracking Database, which describes important FBIS gray literature acquisitions as they are made. NTIS has actively collected and made available many foreign technical reports, including a large collection of South Korean studies. The Library of Congress is working to make more Japanese gray literature available to its users. And DTIC has been working with the British Library Document Supply Centre (BLDSC) to provide access to the latter's excellent collection of conference papers and proceedings. All of these activities could benefit from the broader participation of other Government agencies on the IGLWG.

## **(5) NEEDS FROM THE IT COMMUNITY**

The initial sumMIT conference and this e-journal version of the proceedings are excellent forums for the present paper because many dimensions of gray literature exploitation could benefit greatly from the infusion of Information Technologies (IT). We mention a few areas where gray literature exploitation can be improved in the areas of acquisition, processing, dissemination, and analysis.

Acquisition. The low signal-to-noise ratio of gray literature necessitates that identification and acquisition be separated as separate steps in the collection phase, because a vacuum cleaner approach is not possible with gray literature. Acquisition instead must be demand-driven. Help is needed from the IT to relate availability to requirements, as well as to share awareness of interagency holdings to reduce duplicate acquisitions.

Processing. Much ongoing work in the Government is aimed at improving scanning, machine-aided indexing, optical character recognition, and machine translation, which will benefit all open source exploitation. In gray literature, a tiered processing system needs to be designed so that resources commensurate with the demand for a gray literature product are expended on processing.

Dissemination. This is not generally an issue with gray literature since its real difficulties stem from its acquisition and processing. Still, issues of copyright and electronic document dissemination are important. The size of some technical reports make them harder to transmit or store electronically, while copyright becomes a nightmare as thousands of producers must be tracked down for royalty purposes.

Use. As with processing, much work is ongoing to provide the analyst with tools to analyze digital information. Metrics are needed to evaluate the cost effectiveness of gray literature as a marginal source of intelligence. If it is not providing significant value, then acquisition and processing methodologies must be rethought.

## **CONCLUSIONS**

To improve analyst and information provider efforts to find gray information, we must understand what role gray information can play in solving open source intelligence issues. It is somewhat -- but not entirely -- artificial to think of gray information separately from open source since gray information is a subset of the latter. It is important, however, to know how to use gray information within this broader open source context. One must first exploit what is possible from easier-to-obtain open sources before using possibly marginal sources from the harder-to-obtain gray information domain. But the analyst also must think about how gray information contributes to meeting an information need and in what subject areas it is most advantageous or productive.